TEXAS TECH UNIVERSITY
Natural Science Research Laboratory

Occasional Papers

Museum of Texas Tech University Number 304 29 November 2011

Genetic Diversity in the Transferrin Receptor 1 (*TfR*1) Among Natural Hosts of the North American Arenaviruses

Sheri B. Ayers and Robert D. Bradley

Abstract

For some arenaviruses, cellular entry into a host species may involve the interaction of the host transferrin receptor 1 gene (*TfR*1) and the arenaviral glycoprotein. To examine this scenario, a 42 bp nucleotide region of the *TfR*1 gene, surrounding five conserved amino acid residues (208-212) and the apical domain, was sequenced in host species of North American arenaviruses. The goal was to determine if this region: 1) differed between infected and uninfected individuals of the same host; 2) was polymorphic in a species known to host three genetically divergent strains of a single arenavirus; 3) varied in different rodent species infected with the same arenavirus; and 4) varied between host species and non-host species. Phylogenetic and genetic distance analyses revealed that nucleotides and amino acids were conserved regardless of the host species, arenavirus, or whether a host was infected or uninfected. Results indicated that the *TfR*1 may not be involved in cellular transportation of North American arenaviruses.

Key words: Arenaviridae, arenavirus, *Neotoma*, transferrin receptors, woodrats

Introduction

Viruses of the family Arenaviridae, genus *Arenavirus*, are single-stranded ambisense RNA viruses associated with lymphocytic choriomeningitis and various hemorrhagic fevers in humans (Oldstone 2002; Salvato et al. 2005; Milazzo et al. 2011). The genus *Arenavirus* contains at least 26 species and comprises two genetically and geographically distinct groups (Charrel and de Lamballerie 2003): Lassa-lymphocytic choriomeningitis serocomplex (Old World arenaviruses) and Tacaribe serocomplex (New World arenaviruses). The Tacaribe serocomplex is divided into three South American clades: A (Allpahuayo, Flexal, Paraná, Pichindé, and Pirital); B (Amapari, Cupixi, Guanarito, Junín, Machupo, Sabiá, and Tacaribe); and C (Latino and Oliveros); and a North American clade containing Bear Canyon, Big Brushy Tank, Catarina, Real de Catorce, Skinner Tank, Tamiami, Tonto Creek, Whitewater Arroyo, and other Whitewater Arroyo-like viruses (Bowen et al. 1997; Charrel et al. 2001; Fulhorst et al. 2001; Salvato et al. 2005; Cajimat et al. 2007a, b, 2008, in press; Milazzo et al. 2008, 2010; Inizan et al. 2010).

The Old World arenaviruses are thought to be associated with rodent species in the family Muridae, whereas, New World arenaviruses, with the exception

of the fruit bat *Artibeus jamaicensis* (principal host of the Tacaribe virus - Downs et al. 1963) are associated with members of the rodent family Cricetidae (Cajimat et al. in press). North American arenaviruses, in particular, appear to use members of the rodent subfamily Neotominae as their principal hosts, whereas South American arenaviruses are associated with the subfamily Sigmodontinae.

In order to gain access into a host cell, arenaviruses produce a non-covalently linked surface glycoprotein precursor (GPC) that proteolytically is cleaved into the glycoproteins GP1 and GP2 (Buchmeier et al. 1987; Southern 1996). GP1 mediates arenavirus association with a host cellular receptor and GP2 mediates fusion of the host and viral cellular membranes (Kunz et al. 2005a). Previous studies (Cao et al. 1998; Spiropoulou et al. 2002; Kunz et al. 2005a) indicated that Old World and South American clade C viruses use the host surface cell protein α-dystroglycan as a cellular receptor to interact with the arenaviral GP1. Recently, the surface protein transferrin receptor 1 (*TfR*1) was identified as the receptor for several pathogenic and non-pathogenic clade B viruses (Radoshitzky et al. 2007; Flanagan et al. 2008; Abraham et al. 2009); although, Flanagan et al. (2008) suggested that the non-pathogenic clade B viruses might use other, unknown receptors. In addition, Radoshitzky et al. (2008) identified five amino acid residues (208-212) in the apical domain of the human transferrin receptor gene (h*TfR*1) as the binding site for the GP1 protein. Alteration of only a few of these residues resulted in *TfR*1 supporting the cellular entry of otherwise non-pathogenic arenaviruses (Radoshitzky et al. 2008; Abraham et al. 2009). In particular, the presence of a tyrosine at residue 211 is suspected to affect the efficiency of arenaviral entry into a host cell (Radoshitzky et al. 2008).

To date, no cellular receptor has been identified for North American arenaviruses, although it has been determined they do not use α-dystroglycan (Reignier et al. 2008), as do the clade C viruses. In addition, Reigner et al. (2008) in a series of binding and inhibition assays demonstrated that the White Water Arroyo virus does not require h*TfR*1 to enter cells. However, it remains possible that some North American arenaviruses use *TfR*1 as a receptor. This hypothesis is supported by the fact that North American arenaviruses are phylogenetically most similar to South American clade B viruses (Charrel et al. 2001, 2008; Salvato et al. 2005; Cajimat et al. 2007a, 2008, in press; Milazzo et al. 2008, 2010). In most of those studies, the GPC gene (particularly the GP1 fragment) was one of the primary genetic markers for ascertaining the phylogenetic relationships. If *TfR*1 and GP1 binding is responsible for North American arenavirus entry into host cells, then the two genes should evolve in a similar fashion with nucleotide sequence conservation or evolution in one being mimicked by the other. If *TfR*1 is not the host receptor for North American arenaviruses, then it would not be subjected to selective pressure by arenaviruses and instead would be constrained by its iron-binding function (Richardson and Ponka 1997), resulting in a higher degree of sequence conservation.

Based on recent studies of nucleotide and amino acid sequences of GP1 among North American arenaviruses (Inizan et al. 2010; Cajimat et al. in press), where high levels of genetic divergence were reported (in some cases as high as 39% and 42%, respectively), it would appear that high levels of genetic divergence would be predicted for the *TfR*1 receptor in hosts of North American arenaviruses. Also, given that most New World human pathogens have been associated with the clade B viruses (Flanagan et al. 2008), it is paramount to determine if the *TfR*1 receptor in hosts of North American arenaviruses possess the same conserved amino acid residues as the clade B viruses. Therefore, the objectives of this study were fourfold. First, determine if *TfR*1 nucleotide sequences vary between infected and uninfected individuals of *Neotoma albigula* from populations where the White Water Arroyo arenavirus infection rate ranged from 0 to 40% (Abbott et al. 2004). Second, determine if *TfR*1 receptors are polymorphic in *Neotoma micropus*, a species known to host three genetically divergent strains of the Catarina arenavirus (Fulhorst et al. 2002b). Third, determine if *TfR*1 receptors vary in rodent species (*Neotoma macrotis* and *Peromyscus californicus*) infected with the Bear Canyon arenavirus (Fulhorst et al. 2002a; Cajimat et al. 2007b). Fourth, determine if *TfR*1 sequences vary between hosts (*Neotoma* spp.) and non-hosts (other species of the Neotominae and Sigmodontinae).

METHODS

Sampling and collecting localities.—Specimens were collected from natural populations during previous studies of host-arenavirus relationships (Fulhorst et al. 2001, 2002a; Abbot et al. 2004; Cajimat et al. 2007a, b, 2008, in press; Milazzo et al. 2008, 2010; Inizan et al. 2010); consequently, it was known whether each individual was antibody-positive, antibody-negative, and with which arenavirus it was associated. Voucher specimens and tissue samples previously deposited into the Natural Science Research Laboratory, Museum of Texas Tech University, formed the basis of this study. Specimen numbers, collection localities, virus type, and GenBank accession numbers are listed in the Appendix. For some outgroup and reference samples, DNA sequences were obtained from GenBank.

Twenty-three individuals representing North American principal host species were selected to represent antibody-positive and negative individuals. Efforts were made to include taxa associated with different arenavirus species and divergent strains. Taxa selected included *Neotoma albigula* (n = 4), *N. floridana* (n = 1), *N. fuscipes* (n = 1), *N. leucodon* (n = 3), *N. macrotis* (n = 1), *N. mexicana* (n = 2), *N. micropus* (n = 8), *Peromyscus californicus* (n = 2), and *Sigmodon hispidus* (n = 1). Taxa chosen to represent hosts of South American clade B viruses included *Calomys callosus* (n = 1), *C. musculinus* (n = 1), *Neacomys spinosus* (n= 1), and *Zygodontomys brevicauda* (n = 1). Four individuals were selected from two woodrat species not known to carry an arenavirus, *N. lepida* (n = 3) and *N. stephensi* (n = 1), and 10 individuals were chosen from non-host taxa including *Baiomys taylori* (n = 1), *Cricetulus griseus* (n = 1), *Peromyscus leucopus* (n = 1), *Reithrodontomys fulvescens* (n = 1), *Mus musculus* (n = 1), *Rattus norvegicus* (n = 1), *Equus caballos* (n = 1), *Canis familaris* (n = 1), *Felis catus* (n = 1), *Artibeus jamiacensis* (n = 1), and *Homo sapiens* (n = 1).

RNA isolation and amplification.—To avoid introns and to focus on regions specifically identified by Radoshitzky et al. (2008), mRNA specific to the *TfR*1 gene was examined in a subset of taxa (seven individuals). Total RNA was isolated from frozen liver tissue using the RNeasy RNA isolation kit (Qiagen, Valencia, California). First-strand cDNA copies were generated from mRNA using reverse transcription (Superscript II reverse transcriptase, Invitrogen Life Technologies, Inc, Carlsbad, California) and an oligo (dT) 12-18 primer. A 1,500 base region of the *TfR*1 gene was amplified with primers generated from known sequences of *C. callosus*, *C. musculinus*, and *Z. brevicauda* (Radoshitzky et al. 2008). This nucleotide sequence, consisting of exons 4-17, encoded the ectodomain of the *TfR*1 protein, including the apical domain that interacts with the arenavirus GP1 protein. Also included in this region is the nucleotide sequence that encodes the amino acid loop corresponding to h*TfR*1 208-212, implicated by Radoshitzky et al. (2008) as being important in binding to the arenavirus GP1 protein. The polymerase chain reaction (PCR) method (Saiki et al. 1988) was used to amplify the first-strand cDNA. Primers TfR1-5' (ACAACTATGATGGATCAAGCCAGATCAGCA) and TfR1-3' (ACAACTACATTTAAAACTCATTGTCAATATTCCAAATGTC) were used with the following step-down PCR protocol for primary amplification: 2 cycles of 95°C (30s) denaturing, 60°C (45s) annealing, 72°C (2 min 30s) extension, 2 cycles of 95°C (30s) denaturing, 58°C (45s) annealing, 72°C (2 min 30s) extension, 2 cycles of 95°C (30s) denaturing, 56°C (45s) annealing, 72°C (2 min 30s) extension, 2 cycles of 95°C (30s) denaturing, 53°C (45s) annealing, 72°C (2 min 30s) extension, 27 cycles of 95°C (30s) denaturing, 49°C (45s) annealing, 72°C (2 min 30s) extension, followed by a final extension of 72°C (15 min). Following primary amplification, another PCR protocol was used with primers TfR1-73F (CTGGCTCGGCAAGTAGATGGAGATAA) and TfR1-1917R (GAACTGGTTCAGATCCTTCACAAATGACAG) for nested amplification: 35 cycles of 95°C (30s) denaturing, 48°C (45s) annealing, 72°C (1 min 30s) extension, followed by a final extension of 72°C (15 min).

DNA isolation and amplification.—Based on the nucleotide sequences obtained from the RNA amplifications, a 1,800 base region of the *TfR*1 gene was amplified in the seven selected taxa that targeted the apical domain (regions surrounding amino acids 208-212). The amplified region included: 146 bp of exon 5, intron 5, 108 bp of exon 6 (contains the nucleotide sequence that encodes amino acids 208-212), intron 6,

and 26 bp of exon 7, producing a total coding length of 280 bp. Genomic DNA was isolated from liver or muscle tissue using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, Minnesota). PCR primers were constructed using *Neotoma* sequences generated from RNA (TfR1-440F: AGCTGAGCCA-GAATACAAATAC; TfR1-783R: CCCTGCTCTAA-CAATCACTATAGATCC). The following step-down PCR protocol was used for amplification: 2 cycles of 95°C (30s) denaturing, 60°C (1 min) annealing, 72°C (1 min 40s) extension, 2 cycles of 95°C (30s) denaturing, 58°C (1 min) annealing, 72°C (1 min 40s) extension, 2 cycles of 95°C (30s) denaturing, 56°C (1 min) annealing, 72°C (1 min 40s) extension, 2 cycles of 95°C (30s) denaturing, 53°C (1 min) annealing, 72°C (1 min 40s) extension, 27 cycles of 95°C (30s) denaturing, 50°C (45s) annealing, 72°C (1 min 40s) extension, followed by a final extension of 72°C (15 min).

PCR purification and sequencing.—PCR amplicons were purified using ExoSAP-It (USB Corp., Cleveland, Ohio). All samples were sequenced on an ABI 3100-*Avant* automated sequencer using ABI Prism Big Dye Terminator v.3.1 (Applied Biosystems, Foster City, California). Sequences were proofed and aligned using Sequencher 4.8 (Gene Codes, Ann Arbor, Michigan) and MEGA4 (Tamura et al. 2007). DNA sequences generated in this study were deposited in GenBank.

Data analysis.—For this study, the nucleotide sequence dataset was truncated from the 280 bp region to a 42 bp region containing the apical domain and amino acids 200-212 of the *TfR*1 gene that corresponded to the hypothesized binding sites. Pairwise genetic distance analyses were generated using the nucleotide data under the Jukes-Cantor model of evolution (Jukes and Cantor 1969). To obtain values for comparisons within and among species, the mean of the pairwise comparisons was obtained. Genetic distances were estimated within species and between species, between positive and negative individuals within a species, and between individuals of conspecifics associated with different arenaviruses.

A maximum parsimony analysis was performed using PAUP*4.0 (Swofford 2002) and *H. sapiens* as the outgroup taxon. Nucleotide positions were treated as weighted, unordered, discrete characters with possible states A, C, G, and T; heterozygous sites were designated using the accepted International Union of Biochemistry polymorphic code. Uninformative characters were excluded from analyses and optimal trees were estimated using the heuristic search method with tree bisection-reconnection branch swapping and stepwise addition sequence options. Bootstrap analysis (Felsenstein 1985) with 1,000 iterations was used to evaluate nodal support. In addition, nucleotide sequences were translated into amino acids and analyzed as described previously. Phylogenetically informative sites between residues 200 and 220 that have undergone coding changes were plotted on branches to depict patterns of amino acid evolution.

A Bayesian analysis was conducted using MrBayes software (Huelsenbeck and Ronquist 2001) with *H. sapiens* as the outgroup. MODELTEST (Posada and Crandall 1998) identified the GTR+I+G model as the most appropriate model of evolution. *TfR*1 sequences were partitioned by codon, four Markov chains were run for 10,000,000 generations, and trees sampled every 1,000 generations. The first 10,000 trees were discarded to insure that unstable trees were removed from analyses; remaining sampled trees were inputted into PAUP (Swofford 2002) and a majority-rule consensus tree was constructed. The analysis was repeated four times. Clade probability values ≥ 0.95 indicated nodal support and were included on the phylogenetic tree.

RESULTS

Average distance values (Jukes and Cantor 1969) were estimated for several comparisons and are shown in Table 1. Distances between virus positive and virus negative individuals within a species ranged from 0.41% (*N. leucodon*) to 0.49% (*N. micropus*). The average distance between individuals of *N. micropus* that hosted different arenaviruses was low, ranging from 0.00% (between hosts of the three Catarina virus strains) to 1.10% (between hosts of the Catarina virus and a Whitewater Arroyo-like virus). Species that hosted multiple viruses, *N. albigula* (0.41% between hosts of Big Brushy Tank and Tonto Creek viruses) and

Table 1. Average genetic distances (AGD), based on nucleotide sequences from the TfR1 gene, for selected comparisons of taxa examined in this study. Abbreviations are as follows: Bear Canyon Virus (BCNV); Skinner Tank Virus (SKTV); Tonto Creek Virus (TTCV); and White Water Arroyo Virus (WWAV).

Comparison	AGD
Within Genera (includes virus positive and negative species)	
Neotoma	1.58%
Peromyscus	4.95%
Calomys	4.18%
Within Species (includes virus positive and negative species)	
N. albigula	0.54%
N. lepida	0.69%
N. leucodon	0.55%
N. mexicana	0.00%
N. micropus	0.61%
Between Virus Positive and Virus Negative Individuals within a Species	
N. micropus	0.48%
N. leucodon	0.41%
Between Hosts of BCNV	
N. fuscipes–N. macrotis	0.41%
N. fuscipes–P. californicus	6.60%
N. macrotis–P. californicus	7.04%
Between Hosts of Different Arenaviruses	
N. micropus (3 strains of CTNV)	0.00%
N. micropus (CTNV and WWA-like 1)	1.10%
N. micropus (WWA-like 1 and WWA-like 2)	0.41%
N. albigula (BBTV and TTCV)	0.41%
N. mexicana (SKTV and WWAV)	0.00%

N. mexicana (0.00% between hosts of Skinner Tank and Whitewater Arroyo viruses), depicted low genetic distance values relative to other species. Similarly, individuals of a single species of *Neotoma* that hosted the same virus possessed low values, averaging 0.48%. Average distance values between hosts of the Bear Canyon virus ranged from 0.41% (*N. fuscipes* to *N. macrotis*) to 7.04% (*N. fuscipes* to *P. californicus*). The average genetic distance between species of *Neotoma* was 1.47% and was lower than average distances between species of *Calomys* (4.18%) and species of *Peromyscus* (4.95%).

In the maximum parsimony analysis, a strict consensus tree (Fig. 1) was generated from 2,652 most-parsimonious trees and the resulting topology was characterized by several unresolved clades with few synapomorphies positioned at terminal branches.

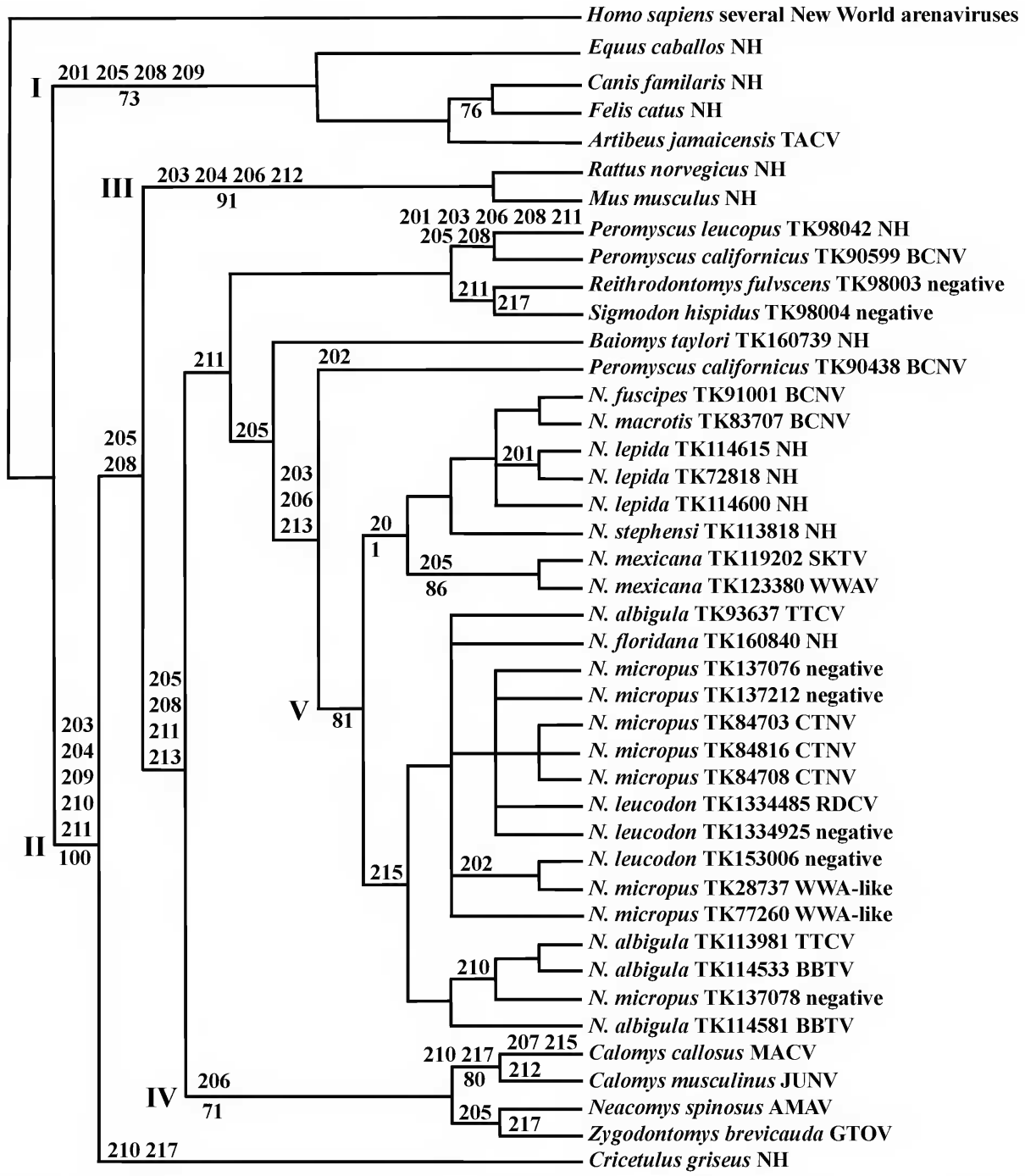

Figure 1. Unweighted maximum parsimony tree obtained from analysis of T*fR*1 nucleotide sequence data. Numbers above branches represent amino acid positions that unite clades. Values below branches are bootstrap support values (only values ≥ 70 are shown). Taxon labels include specimen identification number (TK) and virus status (NH) or virus acronym. Abbreviations for arenaviruses are in the Appendix.

Five major clades were well-supported; one containing individuals of *Equus*, *Canis*, *Felis*, and *Artibeus* (I, bootstrap value = 73), one representing Rodentia (II, bootstrap value = 100), one including the individuals of *Rattus* and *Mus* (III, bootstrap value = 91), one representing the clade containing the South American hosts (IV, bootstrap value = 100), and a clade containing members of the genus *Neotoma* (V, bootstrap value = 81). Within clade IV (hosts for virus clade B), *Calomys callosus* and *C. musculinus*, that possess different arenaviruses, displayed three amino acid substitutions within the 208-212 loop. Few individuals of the same species formed well-supported clades, although the two individuals of *N. mexicana*, one hosting SKTV and one hosting WWAV, formed a monophyletic group. Antibody-positive and negative individuals did not share any consistent amino acid, nor did hosts of different viruses.

The Bayesian analysis produced a tree (Fig. 2) that was similar in topology to that obtained from the parsimony analysis. Four clades were supported with clade probability values ≥ 0.95. Clade I corresponded to the order Rodentia; clade II contained members of the family Muridae (*Mus* and *Rattus*); clade III contained the hosts for the South American arenaviruses; and clade IV contained members of the genus *Neotoma*. Relationships among species and among individuals were characterized by the presence of short branch lengths and the accumulation of few nucleotide substitutions.

A pattern of sequence conservation (G - L Y L), corresponding to the putative binding site (residues 208-212), was conserved among nearly all species of *Neotoma*, with the exception of three individuals that possessed a substitution at amino acid 210 (Table 2). This sequence was present in all woodrats, including *N. lepida* and *N. stephensi*, which based on available data, do not host arenaviruses. This same sequence also was present in species other than *Neotoma*, including North American hosts (*P. californicus*), South American clade B hosts (*Neacomys spinosus* and *Z. brevicauda*) and non-host species (*B. taylori*). Tyrosine-211, which has been suggested to be necessary for cellular entry of arenaviruses (Radoshitzky et al. 2008), was present in all North American hosts (both antibody-positive and negative individuals, with the exception of *S. hispidus*), all South American clade B hosts, and the non-host *B. taylori*.

DISCUSSION

Comparisons of genetic distances (Table 1) revealed low sequence divergence within and between species, with an average distance of 0.48% within species of *Neotoma* and 1.47% between species of *Neotoma*, suggesting a high level of conservation in the *TfR*1 gene. Distance values between positive and negative individuals within a species, as well as hosts of different viruses within a species, were comparable to overall distances between individuals within a species. Average distances between hosts of Bear Canyon virus (*N. macrotis* and *N. fuscipes* to *P. californicus*) were comparable to distances between *Neotoma* and *Peromyscus*. Although *N. micropus* is known to host three divergent strains of Catarina virus in southern Texas (Fulhorst et al. 2002b), the degree of genetic divergence between these three virus strains averaged 12% and was not congruent with the level of genetic divergence of the *TfR*1 genes of their respective hosts (0%). Similarly, hosts of Skinner Tank virus and Whitewater Arroyo virus (individuals of *N. mexicana*) depicted a genetic divergence of 0.00%. In contrast, the genetic distance between *C. callosus* and *C. musculinus* (hosts of clade B viruses) was 4.18%, much higher than values obtained for comparisons within *Neotoma*.

The phylogenetic analyses (Parsimony and Bayesian) depicted clades with low overall support and few synapomorphies (Figs. 1 and 2). The distribution of amino acid substitutions did not appear to determine host specificity. For example, substitution 210 is present in two individuals of *N. albigula*, one that hosts Tonto Creek virus and one that hosts Big Brushy Tank virus. However, two additional individuals of this species do not possess this substitution and yet are associated with these viruses. In contrast, in hosts of Bear Canyon virus, different amino acid substitutions occurred between different hosts of the same virus. For example, there are multiple substitutions that occur along the lineages of *N. micropus* and *N. fuscipes* which do not occur in *P. californicus* and vice versa,

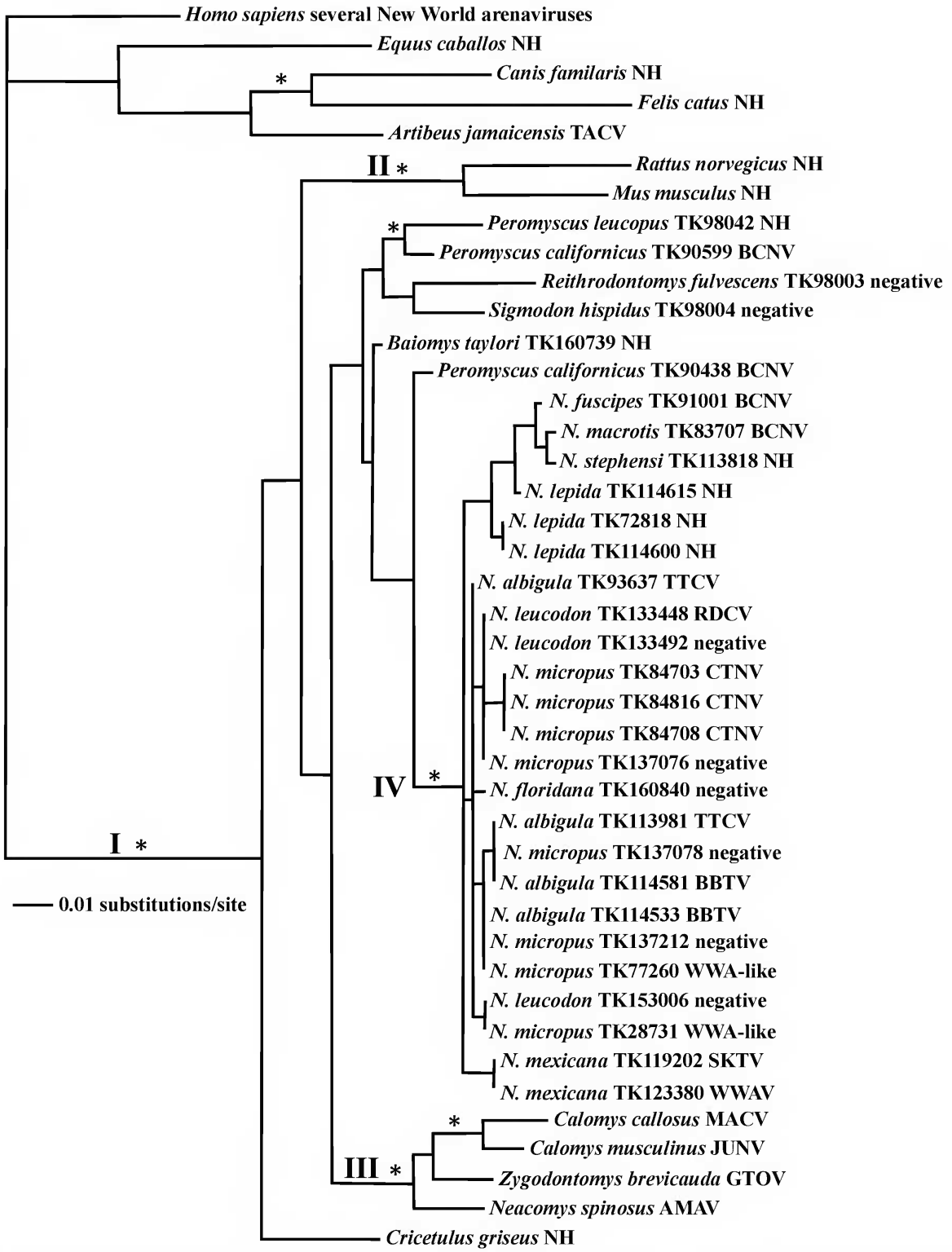

Figure 2. Maximum likelihood tree obtained from analysis of T*fR*1 nucleotide sequence data. Posterior probability values ≥ 0.95 are shown (*) above branches and major clades are depicted by Roman numerals. Taxon labels include specimen identification number (TK) and virus status (NH) or virus acronym. Abbreviations for arenaviruses are in the Appendix.

Table 2. Amino acid differences in the TfR1 gene at residues 204-212 for North American hosts, species not known to host a New World arenavirus, and South American clade B hosts. Sample sizes for each taxon are indicated in parentheses. Abbreviations for amino acids are as follows: A (Alanine), D (Aspartic Acid), E (Glutamic acid), G (Glycine), I (Isoleucine), K (Lysine), L (Leucine), M (Methionine), N (Asparagine), P (Proline), R (Arginine), S (Serine), T (Threonine), V (Valine), and Y (Tyrosine).

	Amino Acid Residue								
Taxon	204	205	206	207	208	209	210	211	212
North American Hosts									
Neotoma spp. (8 species, n = 15)	N	A	S	G	G	-	L	Y	L
Neotoma spp. (2 species, n = 4)	N	A	S	G	G	-	S	Y	L
Neotoma mexicana (n = 2)	N	E	S	G	G	-	L	Y	L
Peromyscus californicus (n = 1)	N	A	S	G	G	-	L	Y	L
Peromyscus californicus (n = 1)	N	E	N	G	A	-	L	Y	L
Sigmodon hispidus (n = 1)	N	D	N	G	G	-	L	N	L
North American Non-hosts									
Baiomys taylori (n = 1)	N	A	N	G	G	-	L	Y	L
Neotoma lepida (n = 3)	N	A	S	G	G	-	L	Y	L
Neotoma stephensi (n = 1)	N	A	S	G	G	-	L	Y	L
Peromyscus leucopus (n = 1)	N	E	S	G	T	-	L	Y	L
Reithrodontomys fulvescens (n = 1)	N	D	N	G	G	-	L	N	L
South American Clade B Hosts									
Calomys callosus (n = 1)	N	A	S	N	G	-	V	Y	L
Calomys musculinus (n = 1)	N	A	S	G	G	-	S	Y	P
Neacomys spinosus (n = 1)	N	S	S	G	G	-	L	Y	L
Zygodontomys brevicauda (n = 1)	N	T	S	G	G	-	L	Y	L
Reference Non-hosts									
Cricetulus griseus (n = 1)	N	V	N	G	D	-	S	D	L
Rattus norvegicus (n = 1)	-	N	S	G	S	N	S	D	P
Mus musculus (n = 1)	Q	S	N	G	N	-	L	D	P
Felis catus (n = 1)	G	T	N	S	G	M	V	Y	L
Canis familiaris (n = 1)	D	M	E	S	D	L	V	Y	L
Equus caballos (n = 1)	N	G	S	G	D	M	I	S	L
Artibeus jamaicensis (n = 1)	A	V	S	S	G	A	G	Y	L
Homo sapiens (n = 1)	D	K	N	G	R	L	V	Y	L

but there are no substitutions shared exclusively by the hosts of Bear Canyon virus. Moreover, there are no amino acid substitutions that differentiate antibody-negative individuals within a species from those that are antibody-positive.

For amino acids 208-212, the sequence G – L Y L was highly conserved (Table 2). Twenty-one of 24 individuals of *Neotoma* possessed this sequence, including hosts of numerous North American viruses (antibody-positive and negative) and species not known to host an arenavirus. Although Radoshitzky et al. (2008) identified this region as a key determinant of host specificity of the clade B viruses, the highly conserved nature of this region in *Neotoma* and *Peromyscus* and homology of this region to *Z. brevicauda* and *Neacomys spinosus* suggests this may not be true for North American viruses.

Despite the lack of genetic variation at amino acids 208-212 of the *TfR*1 gene in North American hosts, specific host-virus relationships are maintained. One possibility is that this particular region of the apical domain is not the only region that interacts with the arenavirus GP1. The tertiary structure of *TfR*1 places this amino acid loop near the α-helix and two β-strands, yet these structures are not continuous along the primary sequence (Lawrence et al. 1999), and it cannot be ruled out that there are other amino acids that may be interacting with GP1. In fact, it has been suggested that a single change at amino acid 348, which is located in an α-helix adjacent to residues 208-212 in the tertiary structure, can interfere with the entry of clade B Machupo and Guanarito viruses (Radoshitzky et al. 2008). There may be a similar mechanism in the North American viruses, in which other residues adjacent to this loop in tertiary structure may affect the conformation of the *TfR*1 protein in such a way to alter host specificity. In addition, amino acid changes elsewhere in the protein, perhaps not in the apical domain or near the GP1 binding site, may affect the tertiary structure indirectly in such a way that impacts the conformation of the binding site.

Although every host species (except *S. hispidus*) contains a tyrosine at position 211, there are species (*N. lepida*, *N. stephensi*, and *B. taylori*) that have a tyrosine at this site, that are not known to host an arenavirus. Although tyrosine-211 may be necessary for arenavirus entry, it alone may not be sufficient. Post-translational modifications at the amino acid loop interact with GP1 and play a role in the efficiency of arenavirus entry. Radoshitzky et al. (2008) identified an N-glycosylation site at residue 205 of *C. callosus*, *C. musculinus*, and *Z. brevicauda*, hosts of Machupo virus, Junín virus, and Guanarito virus, respectively. Removal of this glycosylation motif increases the efficiency of these three viruses in entering the host cells of each of the three host species (Radoshitzky et al. 2008). α-dystroglycan, the receptor used by Old World and clade C arenaviruses, is subject to many post-translational modifications that are necessary for the cellular function of this enzyme (Barresi and Campbell 2006) and coincidentally are necessary for its function as a receptor for arenaviruses (Kunz et al. 2005b). Interestingly, the residue at the 205 site was not conserved in all members of the Cricetidae examined; however, site 204 was conserved. *TfR*1 contains many potential glycosylation and phosphorylation sites that are necessary for proper functioning (Evans and Kemp 1997), and if there are similar glycosylation motifs near or at the binding site, this may be an additional mechanism of maintaining host-specificity and virus entry.

Sigmodon hispidus differed from all other New World hosts by the absence of a tyrosine at residue 211, although this residue also was present in the non-arenavirus associated host *Reithrodontomys fulvescens*. It is important to note that there is genetic differentiation between eastern and western lineages of *S. hispidus* based on AFLP and mitochondrial data (Phillips et al. 2007; Henson and Bradley 2009). The individual examined herein was collected within the range of the western lineage, but the Tamiami virus is known to occur only within the range of the eastern genetic lineage. Future studies should include individuals from southern Florida, within the range of the Tamiami virus, for comparison.

There were no consistent differences in *TfR*1 sequences among comparisons made in this study. This was surprising given the high level of genetic divergences in the GP1 protein (Inizan et al. 2010; Cajimat et al. in press) and suggests that host cell entry may not be determined by *TfR*1 but may involve other unknown proteins as suggested by Flanagan et al. (2008). Alternatively, it may be that differential expression of a specific protein(s) may play a role determining infection and host-specificity (Tayeh et al. 2010).

ACKNOWLEDGMENTS

Thanks to M. S. Corley, H. M. Huynh, N. Ordóñez-Garza, E. K. Roberts, and C. W. Thompson for comments on previous versions of this manuscript. C. F. Fulhorst kindly provided access to arenavirus data generated in his laboratory and provided insight into this study. Portions of this project were funded by a grant from the Texas Tech University Vice-President for Research - Research Enhancement Fund and by the National Institutes of Health (grant AI-41435). Tissue loans were provided by R. J. Baker (Natural Science Research Laboratory, Museum of Texas Tech University).

LITERATURE CITED

Abbott, K. D., M. L. Milazzo, J. Keith, R. D. Bradley, and C. F. Fulhorst. 2004. Epizootiology of arenaviral infections in the white-throated woodrat (Muridae: Sigmodontinae) and other woodrats in Arizona. Journal of Vector Ecology 29:355-364.

Abraham, J., J. A. Kwong, C. G. Albariño, J. G. Lu, S. R. Radoshitzky, J. Salazar-Bravo, M. Farzan, C. F. Spiropoulou, and H. Choe. 2009. Host-species transferrin receptor 1 orthologs are cellular receptors for nonpathogenic New World clade B arenaviruses. PLoS Pathogens 5: e1000358. doi:10.1371/journal.ppat.1000358.

Barresi, R., and K. P. Campbell. 2006. Dystroglycan: from biosynthesis to pathogenesis of human disease. Journal of Cell Science 119:199-207.

Bowen, M. D., C. J. Peters, and S. T. Nichol. 1997. Phylogenetic analysis of the Arenaviridae: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. Molecular Phylogenetics and Evolution 8:301-316.

Buchmeier, M. J., P. J. Southern, B. S. Parekh, M. K. Wooddell, and M. B. Oldstone. 1987. Site-specific antibodies define a cleavage site conserved among arenavirus GP-C. Journal of Virology 61:982-985.

Cajimat, M. N. B., M. L. Milazzo, R. D. Bradley, and C. F. Fulhorst. 2007a. Catarina virus, an arenaviral species principally associated with *Neotoma micropus* (Southern Plains Woodrat) in Texas. American Journal of Tropical Medicine and Hygiene 77:732-736.

Cajimat, M. N. B., M. L. Milazzo, B. D. Hess, M. P. Rood, and C. F. Fulhorst. 2007b. Principal host relationships and evolutionary history of the North American arenaviruses. Virology 367:235-243.

Cajimat, M. N. B., M. L. Milazzo, J. N. Borchert, K. D. Abbott, R. D. Bradley, and C. F. Fulhorst. 2008. Diversity among Tacaribe serocomplex viruses (family Arenaviridae) naturally associated with the Mexican woodrat (*Neotoma mexicana*). Virus Research 133:211-217.

Cajimat, M. N. B., M. L. Milazzo, M. L. Haynie, J. D. Hanson, R. D. Bradley, and C. F. Fulhorst. In press. Diversity and phylogenetic relationships among the North American Tacaribe Serocomplex Viruses (Arenaviridae). Virology.

Cao, W., M. D. Henry, P. Borrow, H. Yamada, J. H. Elder, E. V. Ravkov, S. T. Nichol, R. W. Compans, K. P. Campbell, and M. B. Oldstone. 1998. Identification of α-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. Science 282:2079-2081.

Charrel, R. N., and X. de Lamballerie. 2003. Arenaviruses other than Lassa virus. Antiviral Research 57:89-100.

Charrel, R N., X. de Lamballerie, and S. Emonet. 2008. Phylogeny of the genus *Arenavirus*. Current Opinion in Microbiology 11:362-368.

Charrel, R. N., X. de Lamballerie, and C. F. Fulhorst. 2001. The Whitewater Arroyo Virus: natural evidence for genetic recombination among Tacaribe serocomplex viruses. Virology 283:161-166.

Downs, W. G., C. R. Anderson, I. Spoence, T. H. G. Aitken, and A. H. Hall. 1963. Tacaribe virus, a new agent isolated from *Artibeus* bats and mosquitoes in Trinidad, West Indies. American Journal of Tropical Medicine and Hygiene 12:640-646.

Evans, P., and J. Kemp. 1997. Exon/intron structure of the human transferrin receptor gene. Gene 199:123-131.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Flanagan, M. L., J. Oldenburg, T. Reignier, N. Holt, G. A. Hamilton, V. K. Martin, and P. M. Cannon. 2008. New World clade B arenaviruses can use transferrin receptor 1 (T*fR*1)-dependent and -independent entry pathways, and glycoproteins from human pathogenic strains are associated with the use of *TfR*1. Journal of Virology 82:938-948.

Fulhorst, C. F., M. L. Milazzo, D. S. Carroll, R. N. Charrel, and R. D. Bradley. 2002b. Natural host relationships and genetic diversity of Whitewater Arroyo virus in southern Texas. American Journal of Tropical Medicine and Hygiene 67:114-118.

Fulhorst, C. F., S. G. Bennett, M. L. Milazzo, H. Murray, Jr., J. P. Webb, Jr., M. N. B. Cajimat, and R. D. Bradley. 2002a. Bear Canyon virus: an arenavirus naturally associated with the California mouse (*Peromyscus californicus*). Emerging Infectious Diseases 8:717-721.

Fulhorst, C. F., R. N. Charrel, S. C. Weaver, T. G. Ksiazek, R. D. Bradley, M. L. Milazzo, R. B. Tesh, and M. D. Bowen. 2001. Geographic distribution and genetic diversity of Whitewater Arroyo virus in the southwestern United States. Emerging Infectious Diseases 7:403-407.

Henson, D. D., and R. D. Bradley. 2009. Molecular systematics of the genus *Sigmodon*: results from mitochondrial and nuclear gene sequences. Canadian Journal of Zoology 87:211-220.

Huelsenbeck, J. P. and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Inizán, C. C., M. N. B. Cajimat, M. L. Milazzo, A. Barragán-Gomez, R. D. Bradley, and C. F. Fulhorst. 2010. Genetic evidence for a Tacaribe serocomplex virus, Mexico. Journal of Emerging Infectious Diseases 16:1007-1010.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 in Mammalian Protein Metabolism. (H. N. Munro, ed.). Academic Press, New York, NY.

Kunz, S., J. M. Rojek, M. Perez, C. F. Spiropoulou, and M. B. Oldstone. 2005a. Characterization of the interaction of Lassa fever virus with its cellular receptor α-dystroglycan. Journal of Virology 79:5979-5987.

Kunz, S., J. Rojek, C. Spiropoulou, R. Barresi, K. P. Campbell, and M. B. A. Oldstone. 2005b. Post-translational modification of alpha-dystroglycan, the cellular receptor for arenaviruses by the glycosyltransferase LARGE is critical for virus binding. Journal of Virology 79:14282-14296.

Lawrence, C. M., S. Ray, M. Babyonyshev, R. Galluser, D. W. Borhani, and S. C. Harrison. 1999. Crystal structure of the ectodomain of the human transferrin receptor. Science 286:779-782.

Milazzo, M. L., M. N. B. Cajimat, M, L. Haynie, K. D. Abbott, R. D. Bradley, and C. F. Fulhorst. 2008. Diversity among Tacaribe serocomplex viruses (family Arenaviridae) naturally associated with the white-throated woodrat (*Neotoma albigula*) in the southwestern United States. Vector-Borne and Zoonotic Diseases 8:523-450.

Milazzo, M. L., A. Barragán-Gomez, J. D. Hanson, J. G. Estrada-Franco, E. Arellano, F. X. González-Cózatl, I. Fernández-Salas, F. Ramirez-Aguilar, D. S. Rogers, R. D. Bradley, and C. F. Fulhorst. 2010. Antibodies to Tacaribe serocomplex viruses (family Arenaviridae, genus *Arenavirus*) in cricetid rodents from New Mexico, Texas, and Mexico. Vector-Borne and Zoonotic Diseases 10:629-637.

Milazzo, M. L., G. L. Cambell, and C. F. Fulhorst. 2011. Novel arenavirus infection in humans, United States. Emerging Infectious Diseases 17:1417-1420.

Oldstone, M. B. 2002. Arenaviruses I. The epidemiology, molecular, and cell biology of arenaviruses. Introduction. Current Topics in Microbiology and Immunology 262:v-xii.

Phillips, C. D., C. A. Henard, and R. S. Pfau. 2007. Amplified fragment length polymorphism and mitochondrial DNA analyses reveal patterns of divergence and hybridization in the hispid cotton rat (*Sigmodon hispidus*). Journal of Mammalogy 88:351-359.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Radoshitzky, S. R., J. Abraham, C. F. Spiropoulou, J. H. Kuhn, D. Nguyen, J. Nagel, P. J. Schmidt, J. H. Nunberg, N. C. Andrews, M. Farzan, and H. Choe. 2007. Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. Nature 446:92-96.

Radoshitzky, S. R., J. H. Kuhn, C. F. Spiropoulou, C. G. Albarino, D. P. Nguyen, J. Salazar-Bravo, T. Dorfman, A. S. Lee, E. Wang, S. R. Ross, H. Choe, and M. Farzan. 2008. Receptor determinants of zoonotic transmission of New World hemorrhagic fever arenaviruses. Proceedings of the National Academy of Sciences 105:2664-2669.

Reignier, T., J. Oldenburg, M. L. Flanagan, G. A. Hamilton, V. K. Martin, and P. M. Cannon. 2008. Receptor

use by the Whitewater Arroyo virus glycoprotein. Virology 371:439-446.

Richardson, D. R., and P. Ponka. 1997. The molecular mechanisms of the metabolism and transport of iron in normal and neoplastic cells. Biochimica et Biophysica Acta 1331:1-40.

Saiki, R. K, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermosTable DNA polymerase. Science 239:487-491.

Salvato, M. S., J. C. S. Clegg, M. J. Buchmeier, R. N. Charrel, J. P. Gonzalez, I. S. Lukashevich, C. J. Peters, R. Rico-Hesse, and V. Romanowski. 2005. Family Arenaviridae. Pps. 725-733 in Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. (C. M. Franquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball, eds.). Elsevier Academic Press.

Southern, P. J. 1996. Arenaviruses. Pp. 1505-1519 in Fields Virology. (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, R. Roizman, S. E. Straus, eds.). Lippincott-Raven Publishers, Philadelphia, PA.

Spiropoulou, C. F., S. Kunz, P. E. Rollin, K. P. Cambell, and M. B. Oldstone. 2002. New World arenavirus clade C, but not clade A and B viruses, utilizes α-dystroglycan as its major receptor. Journal of Virology 76:5140-5146.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Tayeh A., C. Tatard, S. Kako-Ouraga, J.-M. Duplantier, and G. Dobigny. 2010. Rodent host cell/Lassa virus interactions: evolution and expression of alpha-Dystroglycan, LARGE-1 and LARGE-2 genes, with special emphasis on the *Mastomys* genus. Infection, Genetics and Evolution 10:2062-2070.

Addresses of authors:

SHERI B. AYERS

Department of Biological Sciences
Texas Tech University
Lubbock, TX 79409-3131

Present Address:
Orchid Cellmark, Inc.
13988 Diplomat Dr., Ste. 100
Farmer's Branch, TX 75234
sheri.ayers@gmail.com

ROBERT D. BRADLEY

Department of Biological Sciences and
Natural Science Research Laboratory, The Museum
Texas Tech University
Lubbock, TX 79409-3131 USA
robert.bradley@ttu.edu

APPENDIX

Specimens in which the *TfR*1gene was sequenced are listed below by GenBank accession number, museum number, and associated arenavirus. Some *TfR*1sequences were obtained from GenBank (sequences generated in previous studies). Abbreviations for New World arenaviruses are as follows: AMAV (Amapari Virus); BCNV (Bear Canyon Virus); BBTV (Big Brushy Tank Virus); CTNV (Catarina Virus); GTOV Guanarito Virus); JUNV (Junin Virus); MACV (Machupo Virus); RDCV (Real de Catorce Virus); SKTV (Skinner Tank Virus); TACV (Tacaribe Virus); TTCV (Tonto Creek Virus); and WWAV (White Water Arroyo Virus). All localities are in the United States unless otherwise noted.

Baiomys taylori.—Texas: Motley County, 1 mi S Flomot (HM044878, TTU109274, negative).

Neotoma albigula.—Arizona: Gila County, White Cow Mine (HM044879, TTU97148, TTCV); Yavapai County, Cherry Creek (HM044880, TTU88387, TTCV); Gila County, Brushy Tank (HM044882, TTU99846, BBTV); and Hackberry Creek (HM044881, TTU99895, BBTV).

Neotoma floridana.—Oklahoma: Blaine County, 2.9 mi S Entrance of Big Bend Rec. Campgrounds (HM044883, TTU109275, negative).

Neotoma fuscipes.—California: Los Angeles County, Zuma Canyon (HM044884, TTU83037, BCNV).

Neotoma lepida.—California: Los Angeles County, West Covina, Galser Wilderness Park (HM044885, TTU88082, negative) and Arizona: Mohave County, Cottonwood Canyon (HM044886, TTU99919, negative; HM044887, TTU99934, negative).

Neotoma leucodon.—MEXICO: San Luis Potosi, 22.8 km N Real de Catorce (HM044888, TTU102969, RDCV); Nuevo Leon, 8.7 km W Doctor Arroyo (HM044889, TTU109270, negative); and Oklahoma: Cimmaron County, Black Mesa State Park (HM044890, TTU109277, WWA-like).

Neotoma macrotis.—California: Riverside County, Rancho Capistrano Ortega Mountains (HM044891, TTU81391, BCNV).

Neotoma mexicana.—Arizona: Coconino County, Skinner Tank (HM044892, TTU100791, SKTV); and Colorado: Larimer County, Sylan Dale Guest Ranch near mouth of Big Thompson Canyon (HM044893, TTU107426, WWAV).

Neotoma micropus.—Oklahoma: Cimmaron County, 1.5 mi S, 3 mi E Kenton (HM044894, TTU43296, WWA-like); New Mexico: Otero County, Fort Bliss (HM044895, TTU79086, WWA-like); and Texas: Dimmitt County, Chaparral Wildlife Management Area (HM044896, TTU80915, CTNV; HM044897, TTU80920, CTNV; HM044898, TTU81029, CTNV).

Neotoma micropus.—Texas: Motley County, 1 mi S Flomot (HM044899, TTU109271, negative; HM044900, TTU109272, negative; HM044901, TTU109273, negative).

Neotoma stephensi.—Arizona: Yavapai County, Pine Flat (HM044902, TTU97595, negative).

Peromyscus californicus.—California: Riverside County, Bear Canyon Trailhead (HM044903, TTU83520, BCNV) and Orange County, 2.4 mi NW El Cariso Ranger Station and Ortega Hwy (HM044904, TTU83562, BCNV).

APPENDIX (CONT.)

Peromyscus leucopus.—Texas: Dimmitt County, Chaparral Wildlife Management Area (HM044905, TTU98086, negative).

Reithrodontomys fulvescens.—Texas: Dimmitt County, Chaparral Wildlife Management Area (HM044906, TTU98091, negative).

Sigmodon hispidus.—Texas: Dimmitt County, Chaparral Wildlife Management Area (HM044907, TTU9892, negative).

*TfR*1 sequences obtained from GenBank. Viruses affiliated with taxon are in parentheses following the GenBank number: *Artibeus jamaicensis* (FJ154605, TACV); *Calomys callosus* (EU164540, MACV); *Calomys musculinus* (EU164541, JUNV); *Canis familaris* (NM001003111, none); *Cricetulus griseus* (L19142, none); *Equus caballos* (DQ284764, none); *Felis catus* (AF276984, none); *Homo sapiens* (NM003234, several); *Mus musculus* (AK088961, none); *Neacomys spinosus* (FJ154604, AMAV); *Rattus norvegicus* (NM022722, none); and *Zygodontomys brevicauda* (EU340259, GTOV).